

Unsupervised Clustering Land Cover Mapping Protocol



Welcome

Introduction

Protocols

Learning Activities

Appendix

Purpose

Use MultiSpec software to produce and classify a land cover type map of the 15 km x 15 km GLOBE Study Site using the MUC System.

Overview

Students use MultiSpec software to group pixels with similar spectral patterns in the Landsat TM data set. Students classify the land cover type of each spectral group using the MUC System. Students use their local expertise of their GLOBE Study Site and their study site measurements to assess the accuracy of their maps.

Student Outcomes

Students gain experience in the use of remotely sensed image data. They gain an understanding of the need for complementary *in situ* data collection. Students create a map and gain a spatial or landscape perspective of their local area.

Science Content

Science and Technology

Technology is essential to science.

Scientific Inquiry

Scientists conduct investigations for a variety of reasons.

Usually there is no one right way to solve a problem.

Geography

Primary

The spatial distribution of population

Maps and satellite-produced images

How to display spatial information on maps and other geographic representations

Middle

Local and global patterns of ecosystems

Physical and human characteristics of places

How to make and use maps and to analyze spatial distributions and patterns

The relative advantages and disadvantages of using maps, satellite images and models to solve geographic problems

Secondary

How to use geographic knowledge, skills and perspectives to analyze problems and make decisions

How to use technologies to represent and interpret Earth's physical and human systems

Land Cover Enrichment

Clustering is a way to separate different land cover classes using spectral patterns.

Assessing the accuracy of a land cover type map

Scientific Inquiry Abilities

Identify questions that can be answered by interpreting satellite images of the GLOBE Study Site.

Interpret satellite imagery in order to answer a question using MultiSpec software, a computer and other appropriate tools and technologies.

Classify land cover and create a land cover type map.

Propose an answer to a question using the land cover type map created.

Evaluate how accurate the land cover type map is using accuracy assessment.

Level

Intermediate and Advanced

Time

Several class periods



Frequency

One time, but may be an iterative process as you progressively investigate more areas within the GLOBE Study Site.

Materials and Tools

Computer

MultiSpec computer software (provided by GLOBE or downloaded from the Web)

512 x 512 pixel Landsat TM image data of the 15 km x 15 km GLOBE Study Site (disk provided by GLOBE)

MUC Field Guide or *MUC Glossary of Terms*

Topographic maps of your area (if available)

Aerial photos of your local area (if available)

Introduction to the MultiSpec Program and the *Unsupervised Clustering Tutorial*

Printed Landsat TM images of the 15 km x 15 km GLOBE Study Site

Getting to Know Your Difference/Error Matrix Field Guide

Preparation

Install the MultiSpec software on the computers students will be using

Install the image on the computers the students will be using

Familiarize yourself with MultiSpec and the *Unsupervised Clustering Tutorial*

Review the MUC System

Accuracy Assessment Tutorial

Prerequisites

Odyssey of the Eyes Learning Activity

Review and practice the *Introduction to the MultiSpec Program* and the *Unsupervised Clustering Tutorial*

Ability to use the *MUC Field Guide* or *MUC Glossary of Terms*

Bird Beak Accuracy Assessment Learning Activity

Unsupervised Clustering Land Cover Mapping Protocol Introduction

Look at the true color satellite image that your teacher has of your GLOBE Study Site. What different colors do you see? What do you think these colors mean? Compare the false color infrared image to the true color. Are the areas that are only one color in the true color image the same size and shape in the false color? What kind of land cover do you think the blue and black colors are? What kind of land cover do you think white or grays are? What is green in the true color image? Find a green area on the true color image; what color is that area in the false color image? What do the different shades of green in the true-color image represent? How is this land cover type represented in the false-color image? Repeat this with the other colors. Try to find your school in the image...it should look like a white or gray set of squares in the center of the image. Are there major roads in the image? What do they look like? Try to make a table (like the one below) that matches each of the different colors and their land cover types.

Work on your own or with a partner. Come up with a list of questions that you would like to answer about the satellite image of your area. From this list, or one that you generate as a class, pick one question that you will try to answer when you create a map from the satellite image. The question can have one or many parts.

Many maps are made from *remotely sensed data*. *Remotely sensed data* is data that is collected from far away using your senses. The protocol GLOBE students use mimics what scientists do. Scientists who make maps from satellite imagery ask questions about their images and use software similar to MultiSpec to answer them. This process uses standard statistical methods to identify map groups, called clusters. *Clusters* are groups of spectrally similar pixels that the computer identifies and groups together according to their patterns of reflectance. The computer gives each cluster a random color. Students must classify the land cover type of each cluster using the MUC System. Some types can be labeled from just student knowledge of the area while other types can be labeled using topographic maps, aerial photographs, etc. and data collected using the *Land Cover Sample Site Protocol*. Scientists also use knowledge that they have learned in the field or from other collected data to label the clusters the computer makes. Once you have many Land Cover Sample Site data, you can perform an accuracy assessment on your map to see how well you classified the land cover in your GLOBE Study Site. There is an *Accuracy Assessment Tutorial* in the *Appendix*, which will walk you through the steps.

The assessed maps made by GLOBE students can be used to help scientists' with their mapping efforts. Some GLOBE student maps can be used to help assess new satellite products (e.g., NASA's EOS satellites).

True-Color Image Color	False-Color Image Color	Type of Land Cover (I think)
Example: White	Example: White	Example: My School - Buildings



Teacher Support

The Measurement

Land cover type mapping is a subjective labeling process where you and your students will have to use your knowledge of your GLOBE Study Site to interpret a computer clustered image. Although this can be a daunting prospect at first, you will find that once you start looking at an image and identifying clusters you know, it does become easier. You will be left to identify fewer and fewer clusters as you continue the process. Just as scientists collect land cover data in the field in order to label their clusters, you and your students should plan to visit sites that cannot be identified using your knowledge, spectral signatures, topographic maps and aerial photos. In those areas, you should carry out the *Land Cover Sample Site Protocol* and report the data to GLOBE. Once your students are proficient with MUC, they can also visit the unknown sites to identify the land cover type on their own. In this case, the site would not be reported to GLOBE but you can use it to label the land cover type map you are creating. When you have completed the map, submit it to GLOBE.

The next step is to see how well your students labeled the image by carrying out an accuracy assessment (by hand or on the GLOBE Website) using additional Land Cover Sample Site data and the *Accuracy Assessment Tutorial*. From here, you can either work on improving the accuracy of your map or see how much change has occurred in your GLOBE Study Site by comparing a 1990's image to a 2000's image. This comparison can be done using the *Change Detection Protocol*.

Teacher Follow-Up: Save the “best” classified clustered image. Use the File menu to save it onto a disk as a TIFF file. If you have a color printer, print copies of your students’ land cover map(s). Report your data to the GLOBE Student Data Archive by e-mailing a file containing the TIFF to globe@ngdc.noaa.gov or by mailing a copy of the TIFF containing your classified, clustered map to:

GLOBE Student Data Archive
NOAA/NGDC E/GC 1
325 Broadway
Boulder CO USA 80305-3328

Supporting Measurements

It may be necessary to carry out the *Land Cover Sample Site Protocol* at some sites where the MUC is unknown. The *Biometry Protocol* measurements may be necessary at these sites.

Student Preparation

Students should spend time discussing what they see in the Landsat images. They should relate their observations to maps, aerial photos, and their own knowledge of the area.

Students must be familiar with the MUC system, and should discuss the type of land covers that are commonly found in their area.

Helpful Hints

- Discuss and identify local examples of land cover types, review topographic maps, and discuss classification before beginning this protocol.
- The Multispec default value of ten clusters is recommended to start with. If this does not appear to adequately represent the land cover types in your image, you may want to increase the number of clusters.
- Remember that this process is iterative. You do not have to label everything at once. Nor will everything be correct. You will improve your map as you get more familiar with your area and take needed ground measurements.
- Have students begin by identifying the most obvious features - usually bodies of water and urban areas - and then progress to more difficult types, such as different types of natural vegetative cover.
- Sometimes cloud shadows resemble lakes and ponds. (See the Beverly, MA image for practice identifying clouds.)
- Use both the true-color and false-color images since some types of land cover will be easier to distinguish in the false-color and others will be easier to distinguish in the true-color.



- Have your students fill in the name of your image on their *Field Guide* before they begin.
- If your students completed the *Manual Interpretation Land Cover Mapping Protocol*, use the same numbers of classes they found in their Student Classified Map or 10 clusters, whichever is greater.
- Hardwood forests may look spectrally similar to actively growing fields.
- The remote sensing image you use may be a few years old. Land cover may have changed since the image was acquired. What you identify on the Landsat TM image may be different than what you see in your ground assessments. In this case, students should work to determine what was on the ground at the time the satellite recorded the image data.
- You will need to field-check areas where you cannot identify the type of land cover. Use the *Land Cover Sample Site Protocol*.

In addition to the color of pixels in a Landsat image, there are a number of keys to help you interpret the land cover types in your image. These keys include shape, size, location, association, and texture. What keys you use depend on the features in your image. Below are some examples of how the keys can be used.

Shape: Agricultural areas tend to have abrupt linear edges and geometric shapes like rectangles and squares. Streams are linear features that can have many bends and curves. Roads frequently have fewer curves than streams.

Size: Major highways and rivers can be distinguished from smaller roads and tributaries.

Topographic or geographic location: If you are in a location with mountains and valleys, forested areas will tend to be on the mountainous areas with steeper slopes, while grassland and agricultural areas will be within the valleys. Since Landsat images are acquired in the morning, hillsides opposite the Sun may be in shadow.

Association: A vegetated area within an urban setting may be a park or cemetery. Wetlands may

be located next to rivers, lakes or estuaries. Commercial centers will be located next to major roads, railroads, or waterways.

Texture: In a false-color image (4-3-2 combination), commercial areas often will be consistently light blue or white. Residential areas, however, may have a speckled appearance of light blue/white and red. The light blue/white indicates buildings and pavement, and the red indicates the grass and trees that may line the streets and surround individual dwellings.

Questions for Further Investigation

Are there distinct land cover classes that this Unsupervised Clustering process did not “resolve,” or separate?

Did the process artificially separate a land cover type into different clusters?

Does increasing the number of clusters (try several different numbers) improve the computer’s ability to resolve these land cover types?

If you had never visited your area but had a Landsat image, what aspects of the local environment would you perceive correctly and what would you not be able to perceive?

What could you do to improve your overall accuracy?

How accurate is your map if someone wanted to find a good place to have a picnic in the woods?

How accurate is your map if you wanted to see how many times you correctly identified a park or playing field?

Which was better, your producer’s or user’s accuracy? Why do you think that is?

How could next year’s class use your data to create a better student classified map?

Unsupervised Clustering

Land Cover Mapping Protocol

Field Guide

Task

Cluster spectrally similar pixels using the MultiSpec software. Label each cluster using your knowledge of the area.

What You Need

- ☐ *Introduction to the MultiSpec Program and Unsupervised Clustering Tutorial*
- ☐ Computer
- ☐ MultiSpec computer software
- ☐ 512 x 512 pixel Landsat TM image data of the 15 km x 15 km GLOBE Study Site
- ☐ *MUC Field Guide* or *MUC Glossary of Terms*

What To Do

1. Start the MultiSpec program on the computer.
2. Open the file containing the Landsat TM image of your GLOBE Study Site. It is labeled _____.
3. Create a new project and select *Cluster* from the **Processor** menu.
4. Select the appropriate number of clusters according to the number of groups you wish to classify. The first time you should use the default of 10 clusters, unless your teacher says otherwise. Provide the system with other information as directed in the *Introduction to the MultiSpec Program* and the *Unsupervised Clustering Tutorial*.
5. Once the image has been clustered, look at the area included in each cluster.
6. Assign a land cover class to each cluster.
 - a. If you know the land cover type of an area, assign a land cover class from the MUC System to the cluster.
 - b. If you do not know the land cover of an area, use the data from a Land Cover Sample Site within the area to assign the land cover class from the MUC System.
 - If there are no land cover sites within the area of a cluster, perform the *Land Cover Sample Site Protocol* for a site within this area.
 - If there are multiple sites within an area, use only one of these sites to make the land cover class assignment and reserve the others for performing an accuracy assessment.
7. Rename each cluster to correspond with its appropriate MUC class.
8. Save the classified clustered image. Use the **File** menu to save it onto a disk as a TIFF file.
9. See your teacher for instructions of how to report your data to GLOBE.



Frequently Asked Questions

1. How many clusters should I use?

For your first attempt at clustering, use the MultiSpec default of 10 clusters.

2. I have used MultiSpec to create 10 clusters. In one of the clusters I know that there are 2 different types of MUC classes. What do I do?

In this case, you should repeat the clustering with additional clusters, for example, this time, use 12. It is also possible that the spectral patterns of the two land cover types are close enough that the software cannot distinguish them.

3. I cannot differentiate roads from commercial areas. What should I do?

Developed areas, MUC 9, are the most difficult to separate in a satellite image. All are composed of minerals and have very similar reflectance patterns. It may not be possible to separate them in a clustering. Sometimes, you can create a separate sub-image of just the urban area and cluster that alone in order to differentiate the different urban types. Do not attempt this unless you are an experienced MultiSpec user.

4. What if I cannot identify a vegetated land cover type from the image to all four levels of MUC?

If you cannot identify the entire MUC class for an area, you will need to visit this area and use the *Land Cover Sample Site Protocol* and any necessary biometry measurements to complete the MUC identification.

5. What if two groups disagree on the MUC value of an area in our image?

If groups cannot agree on the MUC value of an area, you will need to carry out the *Land Cover Sample Site Protocol* and any biometry measurements necessary to settle the disagreement, unless you know someone who lives nearby who can validate the ground cover there.

6. What do we do if there is an area in our image for which nobody knows the MUC value?

Again, the only way to know for certain is to visit the site and collect field data.

7. We have a water body that is not black, but green, or even brown. What does this mean?

In both true color and false-color infrared images, water is normally black. An exception is the very clear water found in parts of the Caribbean. If your water appears brown, green or grey, it usually indicates that there is some material on the surface of the water. This may be plant growth, or suspended sediment that is being transported by the water.



Unsupervised Clustering Land Cover Mapping– Looking at the Data

Are the data reasonable?

After creating a land cover map from your Landsat image, you should determine whether the types of land cover are reasonable and accurate. For instance, if you are located in a mid-latitude temperate climate, does your map include land cover types only found in the equatorial tropical zone? Does it make sense to have land cover types only found in extremely dry desert-like areas? Do you have classes for mountainous areas when you are located in a coastal lowland? Ask yourself questions like these about your own land cover type map. Check the MUC classes and definitions to determine whether the land cover classes in your map make sense for your GLOBE Study Site.

Next, look at where each of these land cover types are located on your map. Using your knowledge of the area and other sources of information for your area like a print-out of your Landsat image, topographic maps and aerial photos (if available), do the locations of the land cover types make sense? If not, which land cover types do not make sense?

After looking at your map and seeing whether it is reasonable, you are now ready to perform a quantitative accuracy assessment. The *Accuracy Assessment Tutorial* provides an example showing how to organize your data and perform the accuracy assessment.

What do scientists look for in these data?

Remote sensing scientists do not have one acceptable overall accuracy percentage to guide them when making a map from satellite imagery. Required accuracy levels depend on the objective of the map. It is very interesting to study the error matrix and observe what land cover classes are being confused with each other. All errors are not equal. In most cases, it would be far worse to label an area as water when it is conifer forest than it would be to label a conifer woodland as a conifer forest. In addition, remote sensing scientists try

to improve their maps using the information gained in the error matrix. These attempts may involve collecting more land cover data to assist in the classification, studying the spectral response patterns of land cover types, and/or applying different classification techniques. Land cover mapping from satellite imagery is often an iterative process and accuracy assessment can take place many times before a final map is achieved. Once scientists have an assessed map, they use it to answer the questions they set out to answer. They may compare the amounts and locations of natural and developed areas, the percentages of land cover types that are important to the community such as agriculture, wetlands, transportation, recreation areas, etc. or specific locations of habitats they are studying.

An Example of Student Inquiry

A group of students in Kyiv, Ukraine were working on creating a land cover map for their GLOBE Study Site. Using MultiSpec, they created an image with 10 clusters. They labeled 8 of the clusters using their knowledge of the types of land cover around their homes and school, and a topographic map of the area. They were uncertain what the MUC classes were for two of the clusters. So they picked out a site for each cluster on the Landsat image. Different students went to each of these sites and performed the *Land Cover Sample Site Protocol*. One site had a MUC code of 811 (Cultivated Land, Agriculture, Row Crop and Pasture). The other site had a MUC code of 92 (Urban, Commercial and Industrial). They used these MUC classes to label the 2 unknown clusters in their image and assumed that all other pixels in these 2 clusters were of the same land cover types.

The students examined the land cover map they created. They discussed how the map looked. They felt fairly confident that the water, closed forest and woodland areas were classified correctly, but questioned whether some of the urban and cultivated areas were correct. Specifically they thought that:

1. roads were included in other MUC 9 classes,
2. bare agricultural fields and commercial areas were not separated into different clusters very well,



Figure LAND-UC-1: Kyiv, Ukraine 15 km x 15 km GLOBE Study Site



Figure LAND-UC-2: Unsupervised Clustered Land Cover Map of Kyiv, Ukraine with Two Unlabeled Classes

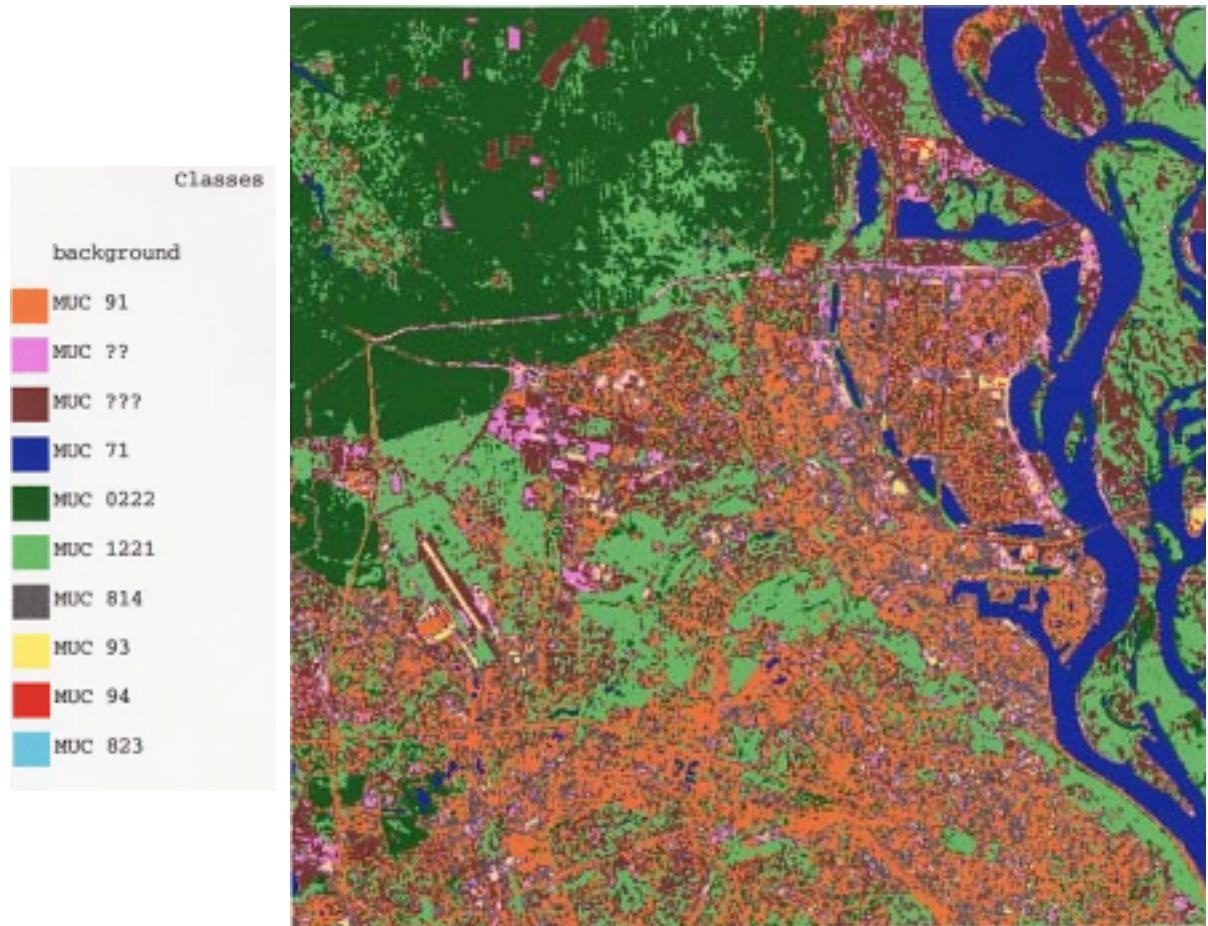
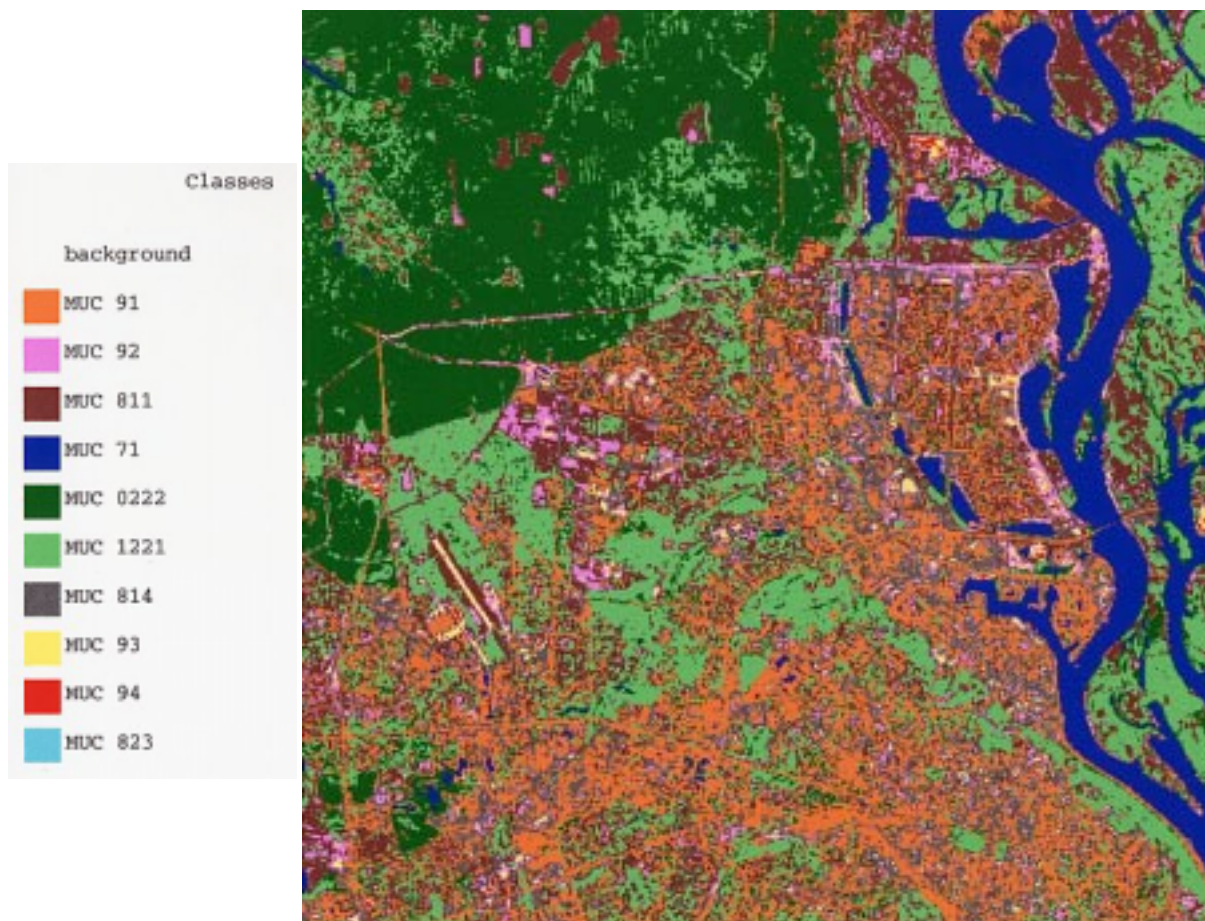


Figure LAND-UC-3: Completed Land Cover Type Map of Kyiv, Ukraine



3. green agricultural fields and non-agricultural areas were not separated into different clusters well.

They decided to see if their hunches were correct.

Earlier in the school year, the group and other classmates collected a number of *Land Cover Sample Site* data. In addition, the class last year collected data for another three sites. Using these

data, they first created a table comparing their classifications of the clustered image and the validation data from the sample sites (Table xx). They remembered NOT to include the 2 sites they used to identify their unknown clusters. These 2 sites were training sites and could not be used for validation data. They had a total of 15 validation sites.

Completed Accuracy Assessment Work Sheet

	Site Name	Student Map Classification Data from GLOBE Study Site	Validation Data from Land Cover Sample Sites	✓	✗
1	River	71	71	✓	
2	Grass by airport	811	824		✗
3	John's Farm	811	811	✓	
4	Downtown Central Square	92	92	✓	
5	Big Woods by Lorenzo's House	0222	0222	✓	
6	Grassland Across River	811	4223		✗
7	LeRoy's Neighborhood	91	91	✓	
8	Highway 66	93	92		✗
9	Natalie's Property	811	824	✓	
10	Airport	93	93	✓	
11	George's Fallow Field	92	811		✗
12	Leonid's Grandma's Neighborhood	92	91		✗
13	Woodland near Janice's House	1222	1222	✓	
14	Forest next to Grandma's House	0222	0222	✓	
15	Mary's Market	91	92		✗

MUC Class List

0222 – Closed Forest, Mainly Deciduous, Cold-Deciduous with Evergreens, With Evergreen Needle-Leaved Trees

1222 – Woodland, Mainly Deciduous, Cold-Deciduous with Evergreens, With Evergreen Needle-Leaved Trees

4223 – Herbaceous Vegetation, Medium Tall Graminoid, With Trees Covering <10%, Trees: Broad-Leaved Deciduous

71 – Open Water, Freshwater

811 – Cultivated Land, Agriculture, Row Crop and Pasture

824 – Cultivated Land, Non-Agriculture, Other Non-Agriculture

91 – Urban, Residential

92 – Urban, Commercial and Industrial

93 – Urban, Transportation

94 – Urban, Other



After creating the table, they made a difference/error matrix. They calculated an overall accuracy of 60%. When they calculated the user's and producer's accuracies, they noticed that the MUC classes for open water, forest and woodland areas were classified correctly, but that many of the developed sites were not. They thought that the data supported their hunches.



		Validation Data									
		MUC 71	MUC 811	MUC 92	MUC 0222	MUC 1222	MUC 91	MUC 93	MUC 824	MUC 4223	Row Totals
Student Map Classification	MUC 71	I									1
	MUC 811		I						II	I	4
	MUC 92		I	I			I				3
	MUC 0222				II						2
	MUC 1222					I					1
	MUC 91			I			I				2
	MUC 93			I				I			2
	MUC 824										0
	MUC 4223										0
	Column Totals	1	2	3	2	1	2	1	2	1	15

They discussed how they might improve their map accuracies. One student thought that it might be a good idea to create a new clustered image using more clusters than the 10 they originally used. She thought that maybe their area had a greater number of different land cover classes than 10 and that with a larger number of clusters, fewer sites would be mis-classified. Another student agreed, but also suggested collecting more *Land Cover Sample Sites*. He observed that many classes only had one validation site. He thought that with more sites, a better accuracy assessment could be made. Everyone in the group agreed and they decided to pursue both suggestions.

Accuracy Assessment Percentages

Overall Accuracy

$$8 \div 15 \times 100 = 53\%$$

User's Accuracies

MUC Class	Calculation	User's Accuracy
71	$1 \div 1 \times 100$	100%
811	$1 \div 4 \times 100$	25%
92	$1 \div 3 \times 100$	33%
0222	$2 \div 2 \times 100$	100%
1222	$1 \div 1 \times 100$	100%
91	$1 \div 2 \times 100$	50%
93	$1 \div 2 \times 100$	50%
824	0	NA
4223	0	NA

Producer's Accuracies

MUC Class	Calculation	Producer's Accuracy
71	$1 \div 1 \times 100$	100%
811	$1 \div 2 \times 100$	50%
92	$1 \div 3 \times 100$	33%
0222	$2 \div 2 \times 100$	100%
1222	$1 \div 1 \times 100$	100%
91	$1 \div 2 \times 100$	50%
93	$1 \div 1 \times 100$	100%
824	$0 \div 2 \times 100$	0%
4223	$0 \div 1 \times 100$	0%